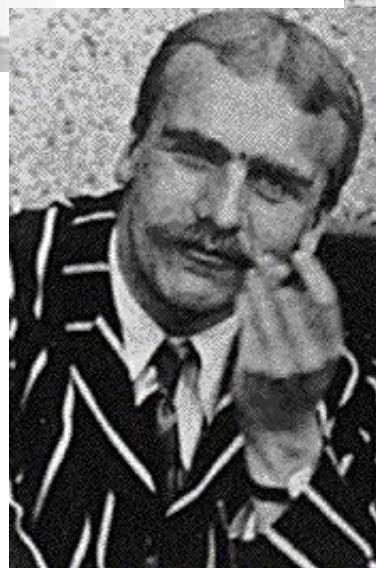
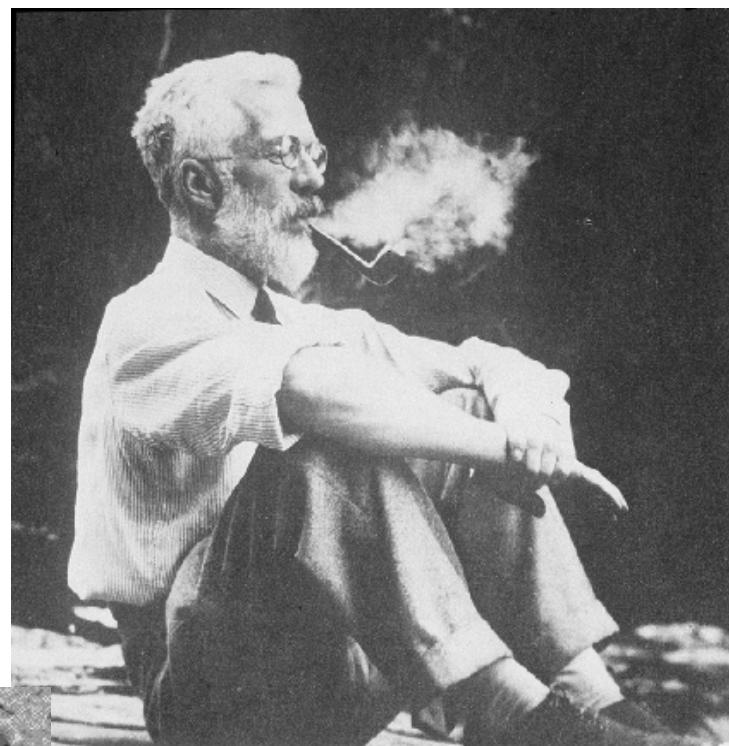
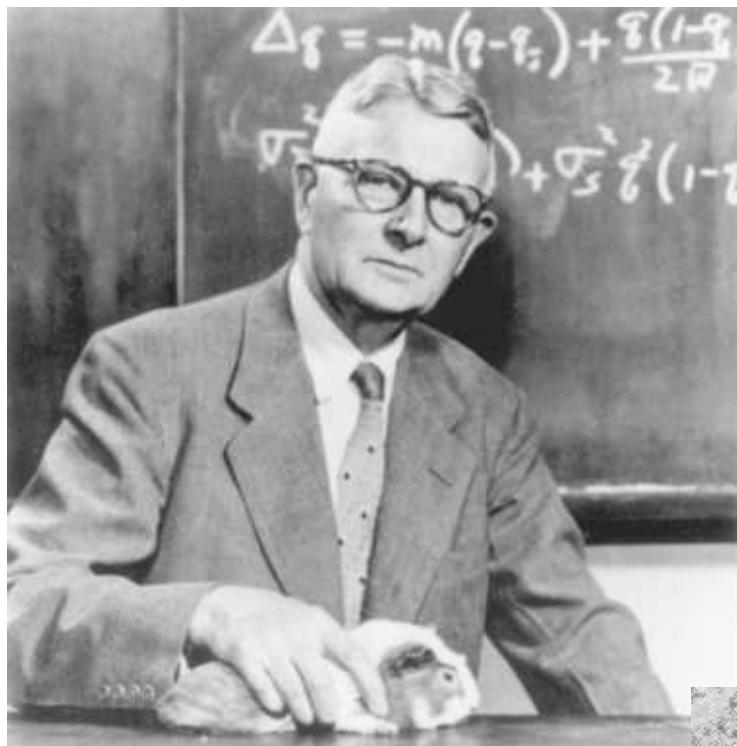
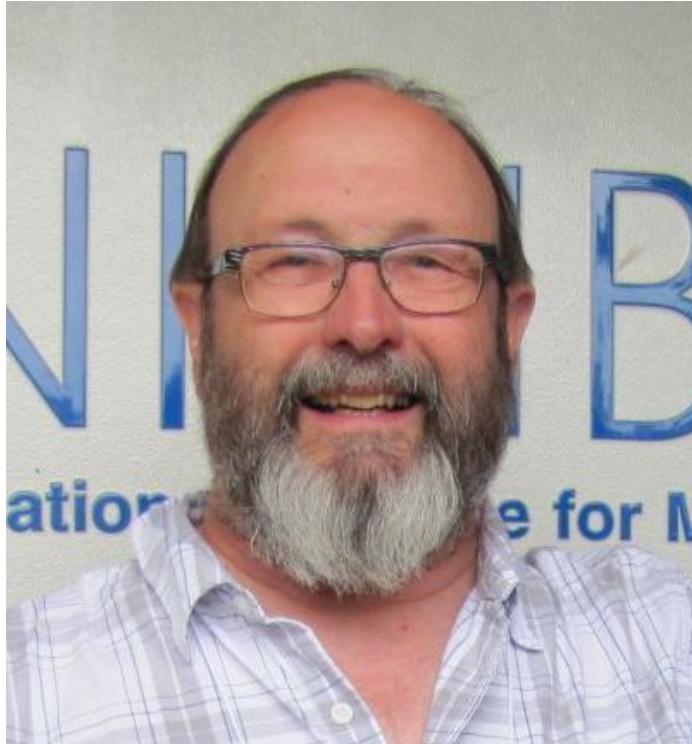


# POP GEN 101





***"We have the same situation in population genomics. People have vast amounts of data and do completely half-ass things with it because they don't know any better. And, I wish there was some way of persuading people that we need to train students in the development and properties of the methods. And that means population genetics."***

**– Joe Felsenstein**

# POPULATION GENETICS

*The study of the distribution of genetic variation within and among populations, and the evolutionary processes that generate this variation.*

*Often, it involves quantifying genetic variation within and among populations, and inferring the evolutionary processes that generated it.*

# **POP GEN 101**

- I. What is genetic variation? (terms and basic principles)**
- II. What determines the amount and distribution of genetic variation?**
  - A. Genetic drift**
  - B. Selection**
  - C. Gene flow**
  - D. Mutation**
- III. Effective population size ( $N_e$ )**
- IV. Neutral vs. adaptive genetic variation**
- V. Why genetics is important in conservation: molecular markers as tools & effect of genetics on fitness**
- VI. Population genomics: What is it? Steps involved? Power and limitations.**

# **POP GEN 101**

**I. What is genetic variation? (terms and basic principles)**

**II. What determines the amount and distribution of genetic variation?**

- A. Genetic drift**
- B. Selection**
- C. Gene flow**
- D. Mutation**

**III. Effective population size (Ne)**

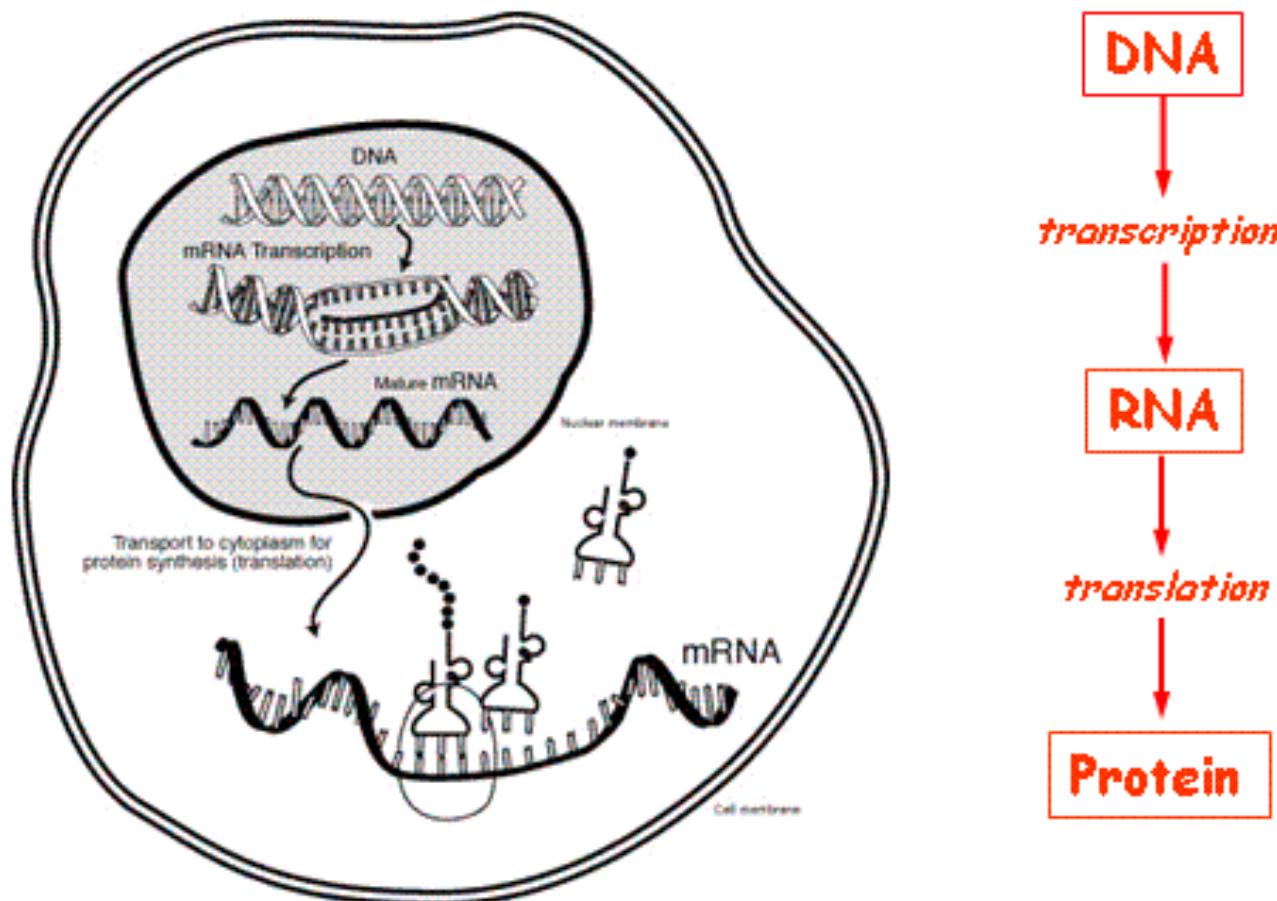
**IV. Neutral vs. adaptive genetic variation**

**V. Why genetics is important in conservation: molecular markers as tools & effect of genetics on fitness**

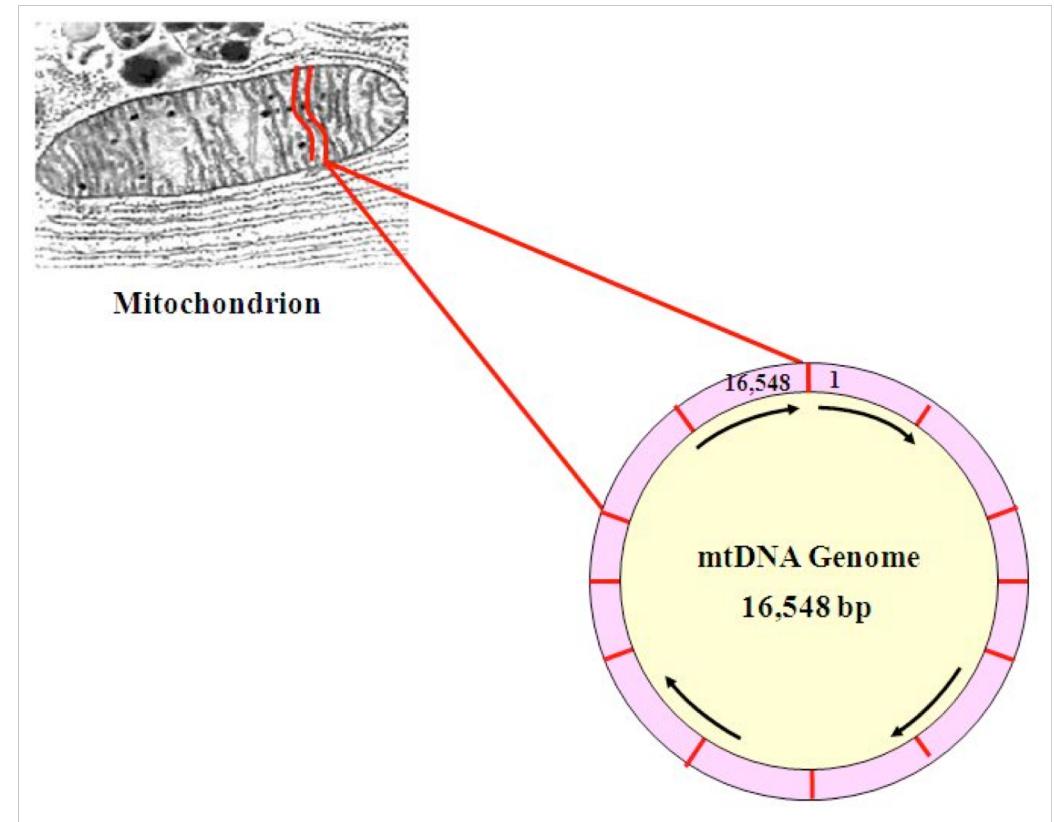
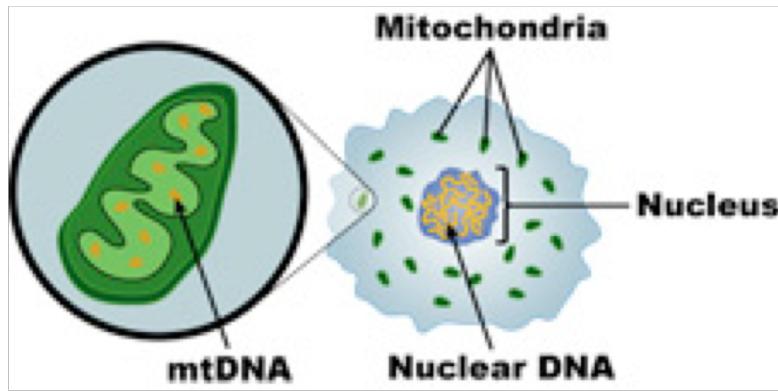
**VI. Population genomics: What is it? Steps involved?  
Power and limitations.**

# Terms and basic principles

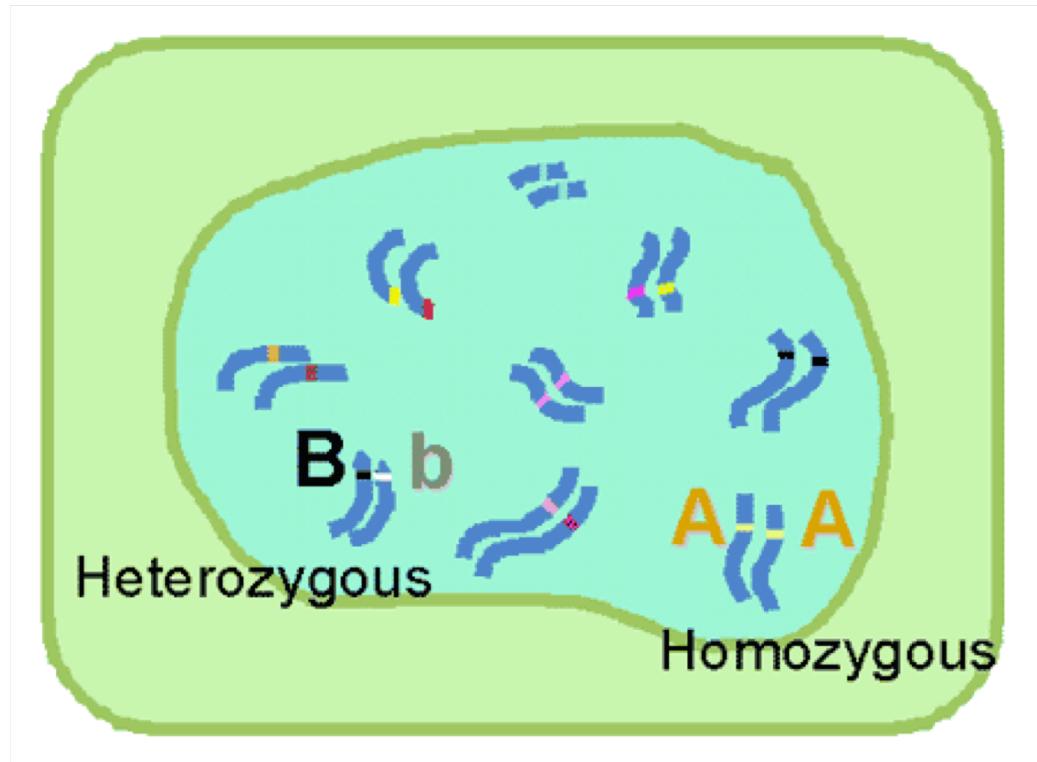
## “Central Dogma” of biology



# mtDNA vs. nDNA



# Terms and basic principles



## Important terms

**Gene**

**Locus (plural: loci)**

**Diploid**

**Allele**

**Heterozygote**

**Homozygote**

# Terms and basic principles

## Hardy-Weinberg proportions

**Figure 5.1** Hardy-Weinberg proportions at a locus with two alleles ( $A$  and  $a$ ) generated by the random union of gametes produced by females and males. The area of each rectangle is proportional to the genotypic frequencies.

		Female gametes (frequency)	
		$A$ ( $p = 0.6$ )	$a$ ( $q = 0.4$ )
Male gametes (frequency)	$A$ ( $p = 0.6$ )	$AA$ ( $p^2 = 0.36$ )	$aA$ ( $qp = 0.24$ )
	$a$ ( $q = 0.4$ )	$aA$ ( $pq = 0.24$ )	$aa$ ( $q^2 = 0.16$ )

## **Terms and basic principles**

### **Hardy-Weinberg equilibrium assumptions**

1. Random mating
2. Large population size
3. No natural selection
4. No gene flow
5. No mutation

### **Outcome of Hardy-Weinberg assumptions**

1. Populations will not evolve (allele and genotype frequencies will remain constant across generations)
2. Genotypic frequencies in H-W proportions after 1 generation of mating

# **POP GEN 101**

- I. What is genetic variation? (terms and basic principles)**
- II. What determines the amount and distribution of genetic variation?**
  - A. Genetic drift**
  - B. Selection**
  - C. Gene flow**
  - D. Mutation**
- III. Effective population size ( $N_e$ )**
- IV. Neutral vs. adaptive genetic variation**
- V. Why genetics is important in conservation: molecular markers as tools & effect of genetics on fitness**
- VI. Population genomics: What is it? Steps involved? Power and limitations.**

# CAUSES OF EVOLUTION: GENETIC DRIFT

## **Hardy-Weinberg equilibrium assumptions**

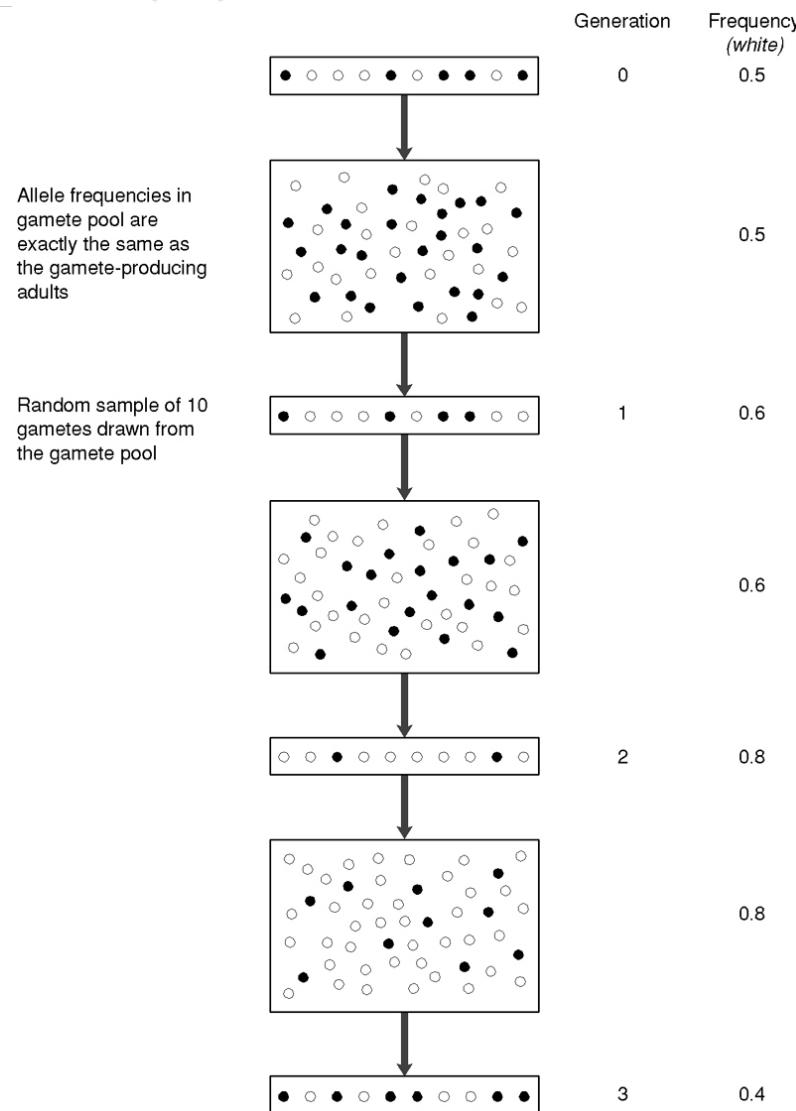
1. Random mating
2. Large population size
3. No natural selection
4. No gene flow
5. No mutation

**What happens when violate assumption of large N?**

- **Genetic drift:** random change in allele frequencies across generations due to finite N
- In other words, finite number of genes transmitted to progeny will be an imperfect sample of allele frequencies in the adults

# CAUSES OF EVOLUTION: GENETIC DRIFT

Figure 6.1 Random sampling of gametes resulting in genetic drift in a population. Allele frequencies in the gamete pools (large boxes) are assumed to reflect exactly the allele frequencies in the adults of the parental generation (small boxes). The allele frequencies fluctuate from generation to generation because the population size is finite ( $N=5$ ). From Grauer and Li (2000).



## 2 results of drift

**1. Allele frequencies will change!**

**2. Genetic variation will be lost!**

$$\Delta h = -1/2N$$

# CAUSES OF EVOLUTION: SELECTION

## **Hardy-Weinberg equilibrium assumptions**

1. Random mating
2. Large population size
3. No natural selection
4. No gene flow
5. No mutation

**What happens when violate assumption of no selection?**

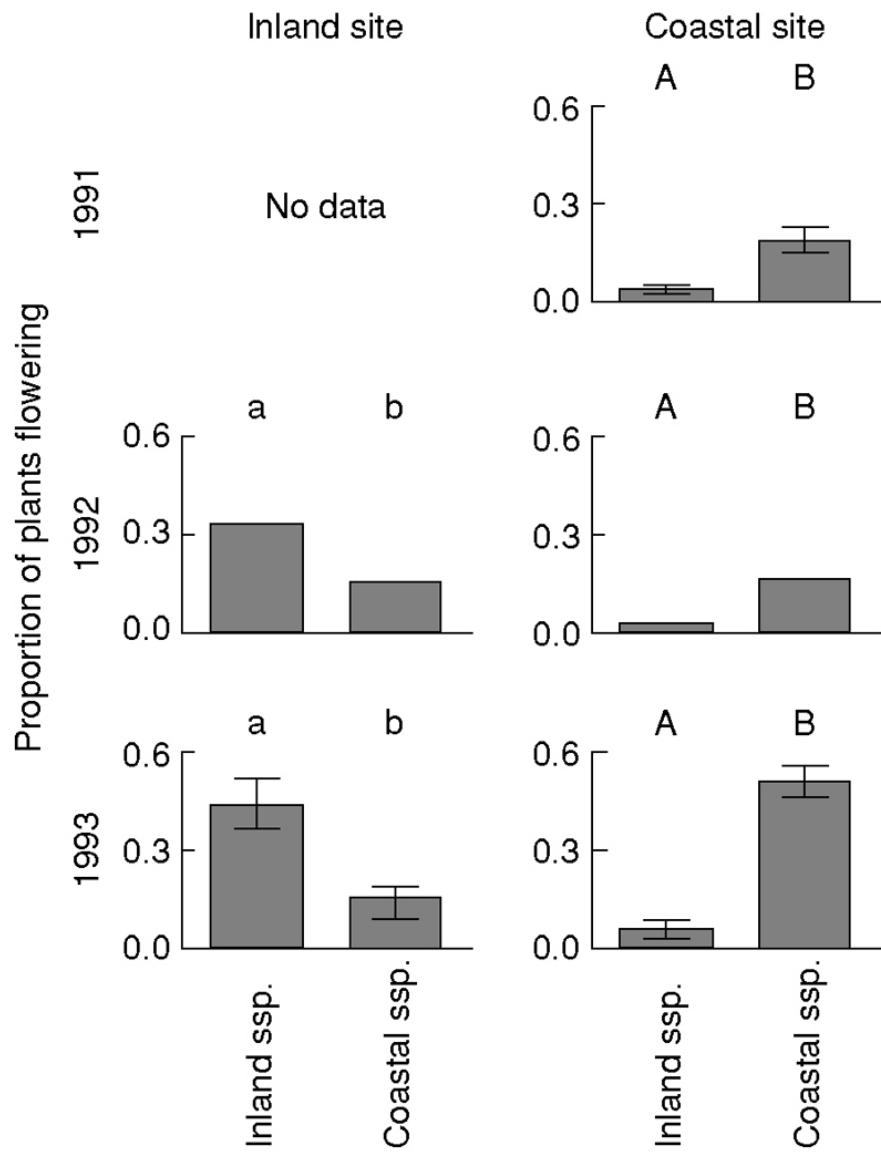
- **Selection:** differential fitness of different alleles & genotypes (many different forms of selection)
- Selection can cause evolution too
- **Local adaptation:** differential fitness of different alleles & genotypes in different environments

# EXAMPLE OF LOCAL ADAPTATION

- Native annual, *Gilia capitata*



- Reciprocal transplant experiment b/n plants on CA coast and inland
- Plants survived to flowering better in their local environment (= local adaptation)



# FITNESS

- **Fitness:** average number of offspring produced by individuals of a particular genotype
  - Effect of natural selection on genotypes measured by fitness
  - Calculated as product of viability and fertility

<u>Genotype</u>	<u>Viability</u>	<u>Fertility</u>	<u>Fitness</u>
AA	$v_{11}$	$f_{11}$	$(v_{11})(f_{11})=w_{11}$
Aa	$v_{12}$	$f_{12}$	$(v_{12})(f_{12})=w_{12}$
Aa	$v_{22}$	$f_{22}$	$(v_{22})(f_{22})=w_{22}$

## **DIRECTIONAL SELECTION**

- Occurs when one allele is always at a selective advantage
- Allele under selection may be dominant, recessive, or intermediate
  - Dominant       $w_{11} = w_{12} > w_{22}$
  - Intermediate     $w_{11} > w_{12} > w_{22}$
  - Recessive         $w_{11} > w_{12} = w_{22}$
- Advantageous allele will increase in frequency and will ultimately be fixed under all three modes of natural selection
- However, rate of change in allele freq depends on mode of dominance relationships

## **HETEROZYGOUS ADVANTAGE**

- Occurs when heterozygote has the greatest fitness:  
 $w_{11} < w_{12} > w_{22}$
- Will maintain both alleles in the population

# CAUSES OF EVOLUTION: GENE FLOW

## **Hardy-Weinberg equilibrium assumptions**

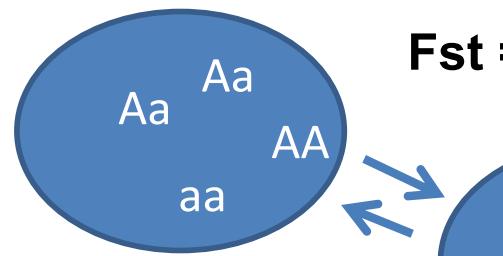
1. Random mating
2. Large population size
3. No natural selection
4. No gene flow
5. No mutation

**What happens when violate assumption of no gene flow?**

- **Gene flow:** movement of alleles from one population to another via dispersal (also called “migration” in pop. genetics)
- Gene flow reduces differences among populations and increases GV within pops.

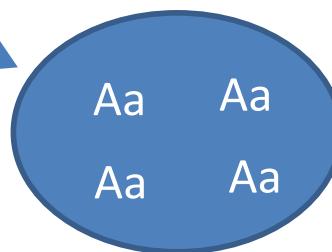
# POPULATION STRUCTURE

**Genetic variation within and among populations**  
>> Fst a measure of genetic differentiation



$$\begin{aligned}f(A) &= 0.5 \\f(a) &= 0.5 \\H &= 0.5\end{aligned}$$

$$Fst = 0.00$$

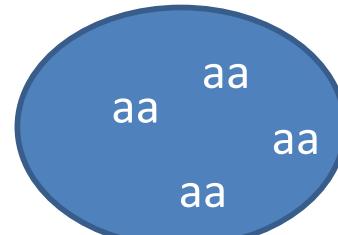


$$\begin{aligned}f(A) &= 0.5 \\f(a) &= 0.5 \\H &= 1.0\end{aligned}$$

$$Fst = 1.00$$

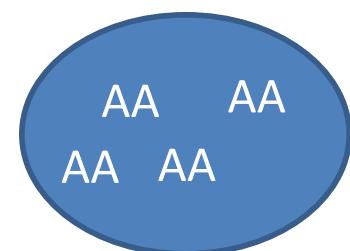
Pop 1

$$\begin{aligned}f(A) &= 0.0 \\f(a) &= 1.0 \\H &= 0.0\end{aligned}$$



Pop 2

$$\begin{aligned}f(A) &= 1.0 \\f(a) &= 0.0 \\H &= 0.00\end{aligned}$$



# CAUSES OF EVOLUTION: GENE FLOW

## **Hardy-Weinberg equilibrium assumptions**

1. Random mating
2. Large population size
3. No natural selection
4. No gene flow
5. No mutation

What happens when violate assumption of no mutation?

- **Mutation:** formation of new alleles caused by changes in DNA
- Ultimate source of genetic variation, increasing within population genetic variation

# MOLECULAR MUTATIONS

## ➤ Types of molecular mutations:

1. Substitutions: replacement of one nucleotide with another
2. Recombinations: exchange of a sequence from one homologous chromosome to the other
3. Deletions: Loss of one or more nucleotides
4. Insertions: Addition of one or more nucleotides
5. Inversions: Rotations by 180° of a double-stranded DNA segment of two or more bp's

# MUTATION

➤ ***How common are mutations?!?!***

▪ **Per locus level:** RARE

- E.g., Few per billion base pairs per generation

▪ **Genome level:** COMMON

- Most species' genomes contain billions of nucleotides
- Most species' genomes contain hundreds of mutations (Lynch et al. 1999)

# **SUMMARY OF EFFECTS OF 4 EVOLUTIONARY FORCES ON GENETIC VARIATION**

<u>FORCE</u>	<u>Effect on genetic variation</u>	
	<u>WITHIN populations</u>	<u>AMONG populations</u>
Genetic drift	↓	↑
Directional selection	↓	↑
Heterozygous advantage	↑	↓
Gene flow	↑	↓
Mutation	↑	↑↓?

# **POP GEN 101**

- I. What is genetic variation? (terms and basic principles)
- II. What determines the amount and distribution of genetic variation?
  - A. Genetic drift
  - B. Selection
  - C. Gene flow
  - D. Mutation
- III. Effective population size ( $N_e$ )
- IV. Neutral vs. adaptive genetic variation
- V. Why genetics is important in conservation: molecular markers as tools & effect of genetics on fitness
- VI. Population genomics: What is it? Steps involved? Power and limitations.

## **EFFECTIVE POPULATION SIZE (Ne)**

- If know population size, can predict rate of loss of heterozygosity

$$\Delta h = -1/2N$$

- BUT... N in above equation not “census population size” (Nc), which is the total number in population
- Instead, N above is the “effective population size” (Ne), which can be thought of as the number of individuals contributing their genes to the next generation
- Many factors affect how much lower Ne is than Nc:
  - Unequal sex ratio
  - Variance in number of progeny
  - Fluctuating population size

## **EFFECTIVE POPUALTION SIZE (Ne)**

- Consider extreme case of one male mating with 100 females



- In this case, all progeny will be half sibs even though total N = 101
  - Can estimate  $N_e$  as:
- $$N_e = \frac{4N_f N_m}{N_f + N_m}$$
- In this case,  $N_e = 4!$

# **POP GEN 101**

- I. What is genetic variation? (terms and basic principles)
- II. What determines the amount and distribution of genetic variation?
  - A. Genetic drift
  - B. Selection
  - C. Gene flow
  - D. Mutation
- III. Effective population size ( $N_e$ )
- IV. Neutral vs. adaptive genetic variation
- V. Why genetics is important in conservation: molecular markers as tools & effect of genetics on fitness
- VI. Population genomics: What is it? Steps involved? Power and limitations.

# ADAPTIVE VS. NEUTRAL GENETIC VARIATION

- Most of your DNA is “junk”!!!

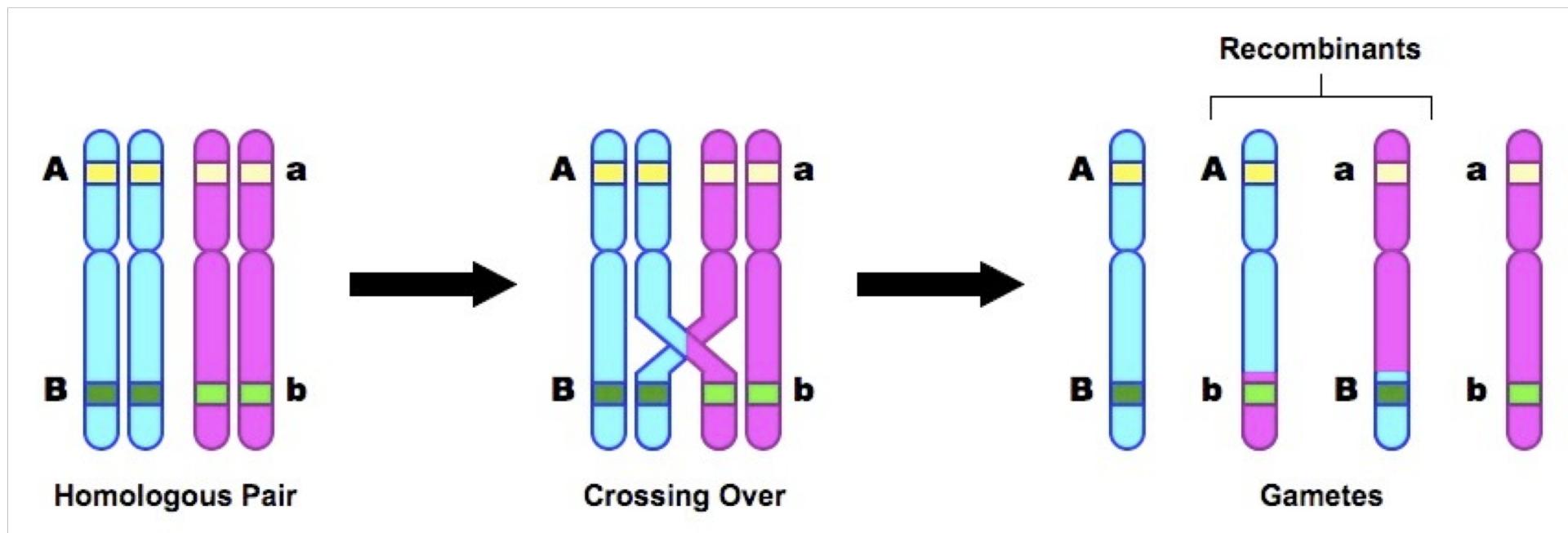


- <2% of the human genome is “coding” (transcribed into mRNA and then translated into proteins)
- Much of the non-coding portion of the genome is “neutral” (b/c not functional, selection doesn’t act on it)
- “Neutral” genetic variation → selection doesn’t act on it (only drift, gene flow, and mutation)
- “Adaptive” genetic variation → selection DOES act on it (in addition to drift, gene flow, and mutation)

# ADAPTIVE VS. NEUTRAL GENETIC VARIATION

➤ How can different evolutionary processes act on different parts of the same genome?!?

➤ Different chromosomes and *recombination*!



# **POP GEN 101**

- I. What is genetic variation? (terms and basic principles)
- II. What determines the amount and distribution of genetic variation?
  - A. Genetic drift
  - B. Selection
  - C. Gene flow
  - D. Mutation
- III. Effective population size ( $N_e$ )
- IV. Neutral vs. adaptive genetic variation
- V. Why genetics is important in conservation: molecular markers as tools & effect of genetics on fitness
- VI. Population genomics: What is it? Steps involved? Power and limitations.

# **WHY GENETICS IS IMPORTANT IN CONSERVATION**

**1. “Molecular markers” (i.e., loci) can be used to infer demographic parameters, evolutionary history, and local adaptation**

- ***Neutral loci:*** Can be used to infer demographic parameters (e.g.,  $N_e$ , gene flow) and evolutionary history (phylogenetic trees)

$$F_{ST} = 1 / (4Nm + 1)$$

- ***Loci under selection:*** Can be used to infer patterns of local adaptation

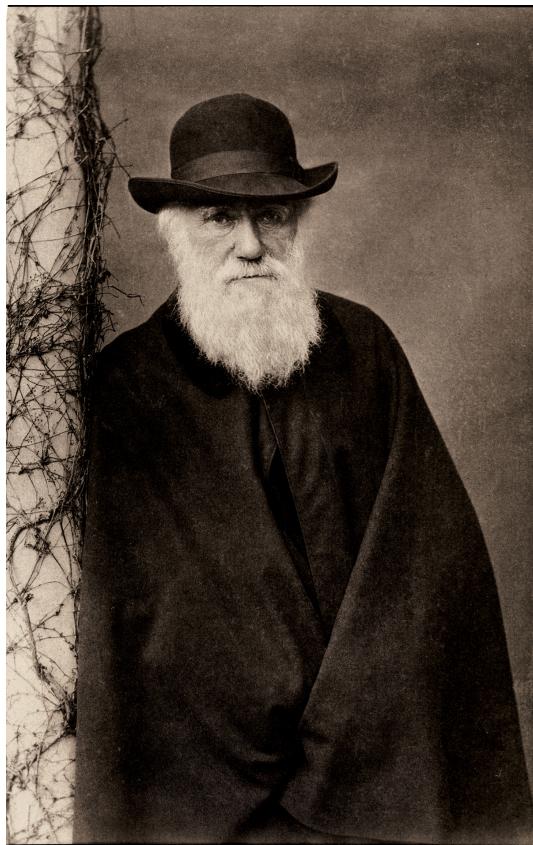
# **WHY GENETICS IS IMPORTANT IN CONSERVATION**

## **2. Genetic variation affects individual fitness, population persistence, and the capacity to adapt**

- ***Inbreeding depression:*** Reduction in fitness due to mating among close relatives
  
- ***Adaptive potential:*** The rate of adaptive evolution is proportional to the amount of genetic variation

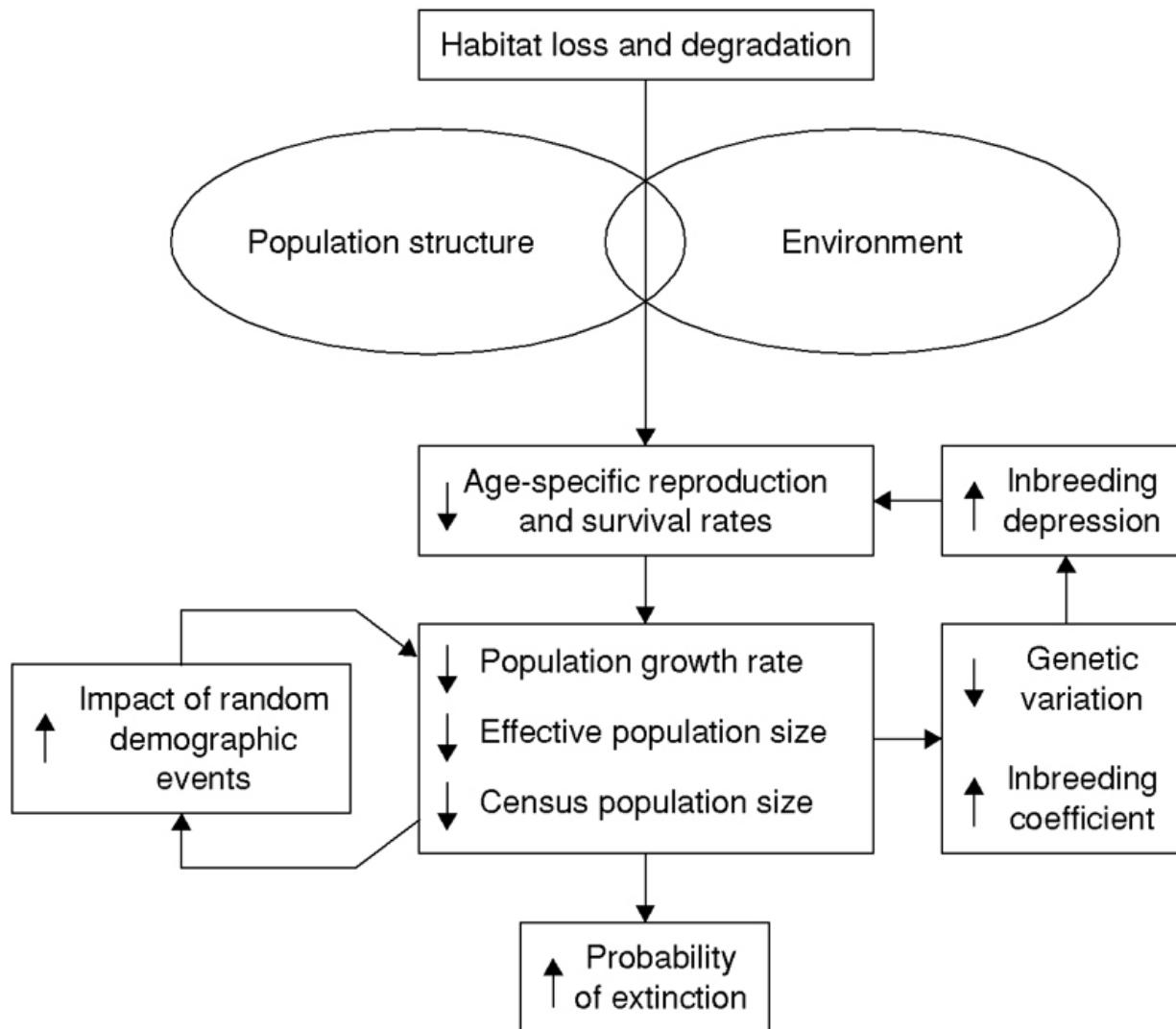
# WHY GENETICS IS IMPORTANT IN CONSERVATION

*“As some of our British parks are ancient, it occurred to me that there must have been long-continued close interbreeding with the fallow-deer (*Cervus dama*) kept in them; but on inquiry I find that it is a common practice to infuse new blood by procuring bucks from other parks.” **Charles Darwin (1896)***



# WHY GENETICS IS IMPORTANT IN CONSERVATION

## ➤ Extinction vortex:



# **POP GEN 101**

- I. What is genetic variation? (terms and basic principles)
- II. What determines the amount and distribution of genetic variation?
  - A. Genetic drift
  - B. Selection
  - C. Gene flow
  - D. Mutation
- III. Effective population size ( $N_e$ )
- IV. Neutral vs. adaptive genetic variation
- V. Why genetics is important in conservation: molecular markers as tools & effect of genetics on fitness
- VI. Population genomics: What is it? Steps involved? Power and limitations.**

# POPULATION GENOMICS

## ➤ ***DEFINITION:***

- Black et al. (2001): “simultaneous sampling of numerous variable loci within a genome and the inference of locus-specific effects from the sample distributions.”
- Luikart et al. (2003): “simultaneous study of numerous loci or genome regions to better understand the roles of evolutionary processes (such as mutation, random genetic drift, gene flow and natural selection) that influence variation across genomes and populations.”

# POPULATION GENOMICS

## ➤ ***DEFINITION:***

- Allendorf & Luikart (2007): “refers to the study of many DNA markers (e.g., mapped markers and coding genes) in many individuals from different populations.”

# POPULATION GENOMICS

## ➤ **GOALS:**

- Identify outlier loci
- Test for signature of selection
- Identify loci under selection
- Infer evolutionary or demographic parameters without bias caused by loci under selection

# POPULATION GENOMICS

## ➤ *RELATIONSHIP W/ POPULATION GENETICS*

- Population genomics just “population genetics writ large”?
- Yes?: Long-recognized that analyzing only a few loci or only one class of markers can provide an incomplete or biased view of genome and of population history or relationships

# POPULATION GENOMICS

## ➤ *RELATIONSHIP W/ POPULATION GENETICS*

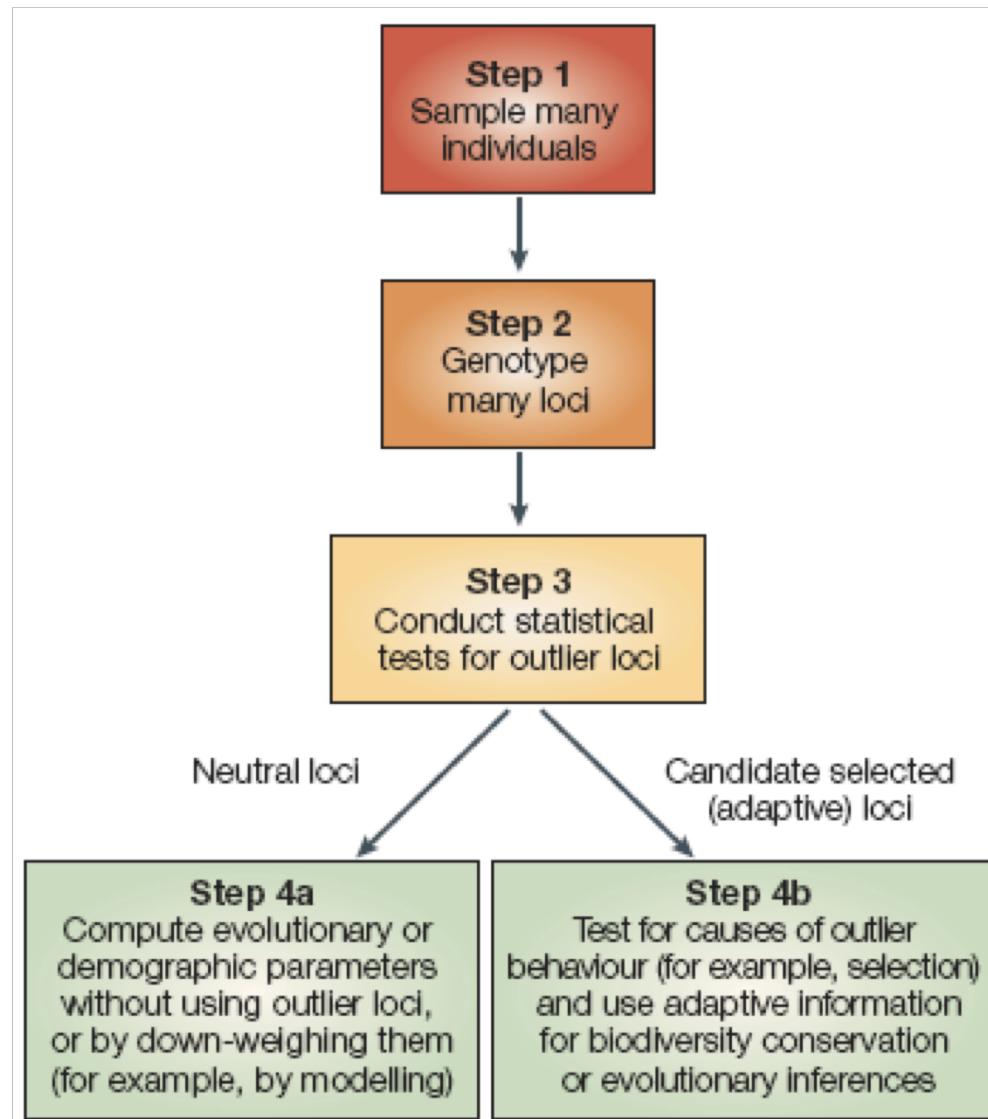
- Population genomics just “population genetics writ large”?
- No?: Only now is it possible to genotype vast numbers of marker loci (“genome typing”) in many individuals and populations of non-model organisms. Also, new software available to ID outlier loci.

# POPULATION GENOMICS

## ➤ *RELATIONSHIP W/ POPULATION GENETICS*

- Thus although population genomics goals not new, we're finally in a position to ID outliers and make evolutionary inferences based on HUGE numbers of loci

# STEPS OF POPULATION GENOMICS



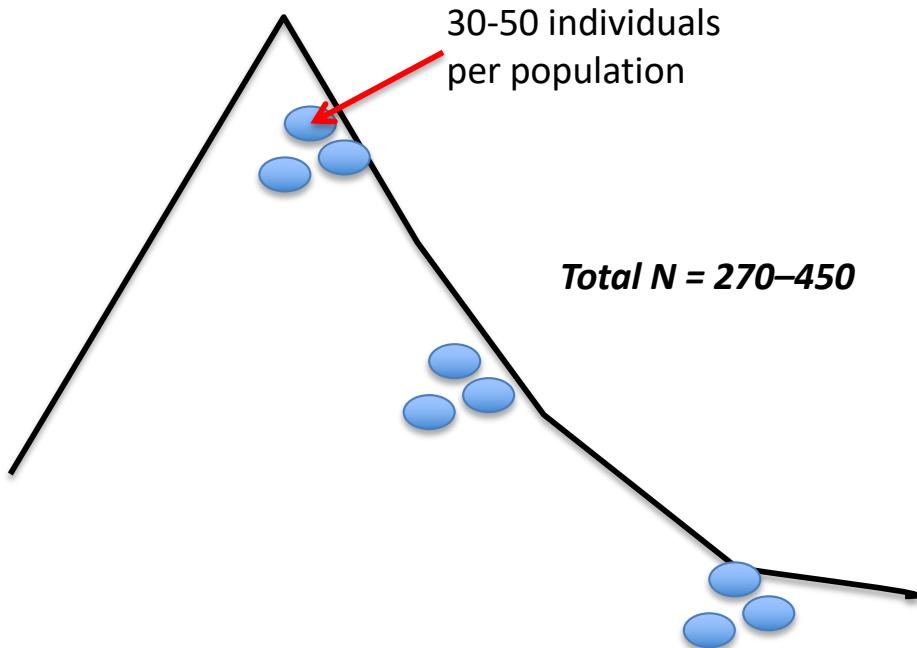
# STEP 1: Sample many individuals

- First step is to sample dozen or hundreds of individuals from one or more populations
- Specific sampling design and sample sizes depends on question:
  - Estimating  $N_e$ : 30-50 individuals from a single population might be adequate
  - Estimating adaptive genetic differentiation: 100s of individuals across a geographic selection gradient

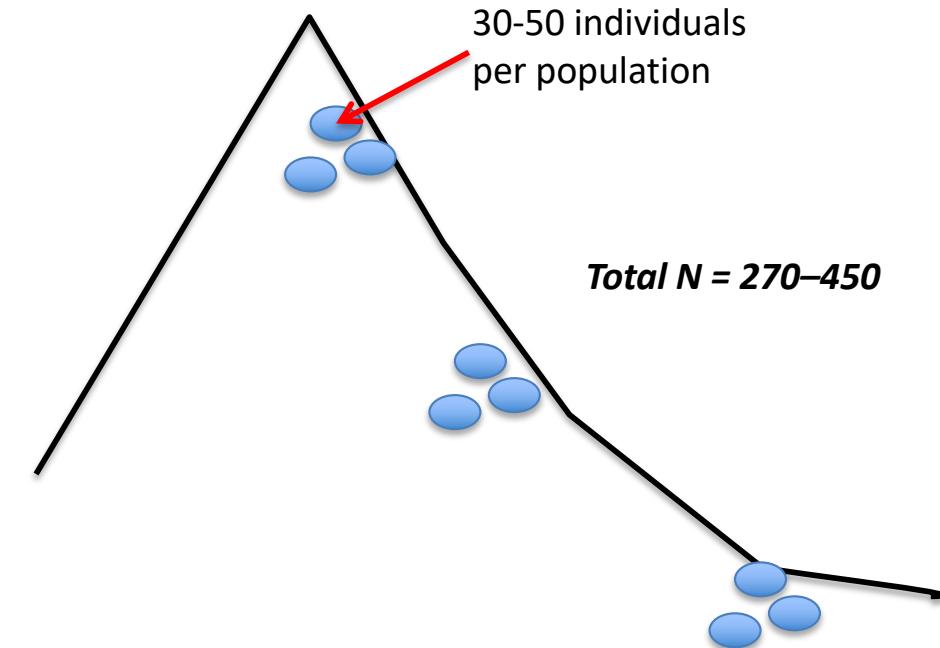
# STEP 1: Sample many individuals

➤ Example of sampling design to test for divergent selection across elevational gradient:

ELEVATIONAL TRANSECT 1



ELEVATIONAL TRANSECT 2



# STEP 2: Genotype many loci

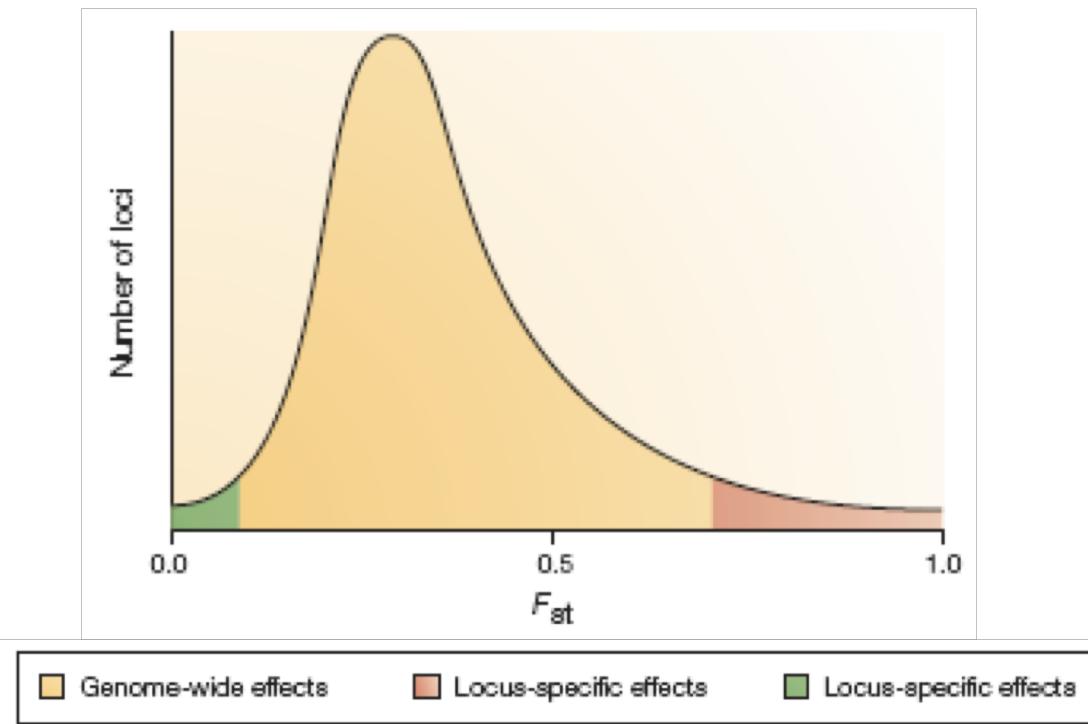
- Second step is to genotype many loci (tens to hundreds):
  - Neutral loci:
    - Prerequisite for accurately inferring demographic history
    - Also necessary for providing neutral baseline to test for statistical outlier loci (potentially under selection)
  - Candidate or functional genes:
    - If don't have candidate loci (i.e., non-models), gene-rich regions can be screened

# STEP 3: Conduct statistical tests for outlier loci

- Third step is to identify outlier loci using statistical tests
- Various types of aberrant behavior that may be caused by selection:
  - Exceptionally high or low  $F_{st}$
  - Excess or deficit of heterozygous genotypes ( $F_{is}$  outliers, where  $F_{is}$  is a index of deviation from HW prop.)
  - Excess of low frequency alleles (homozygosity excess)
  - Deficit of low frequency alleles (heterozygosity excess)

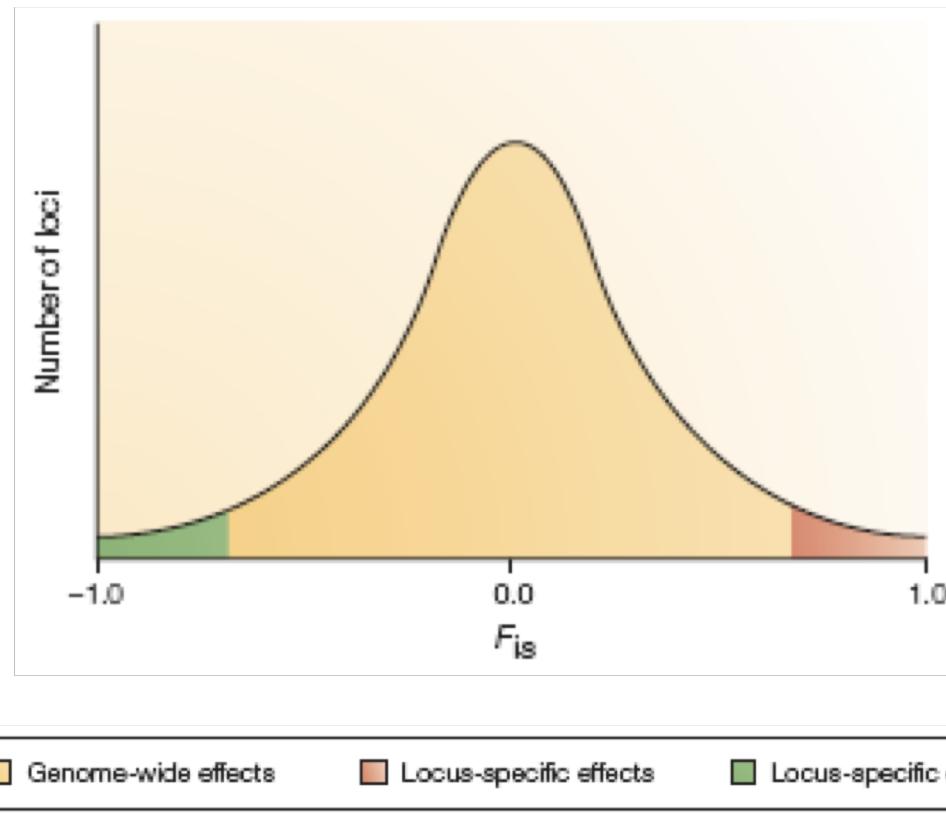
# STEP 3: Conduct statistical tests for outlier loci

- Example of loci that are  $F_{st}$  outliers:
  - What type of selection predicted to cause extremely high  $F_{st}$  outliers?
  - What type of selection predicted to cause extremely low  $F_{st}$  outliers?



# STEP 3: Conduct statistical tests for outlier loci

- Example of loci that are  $F_{is}$  outliers:

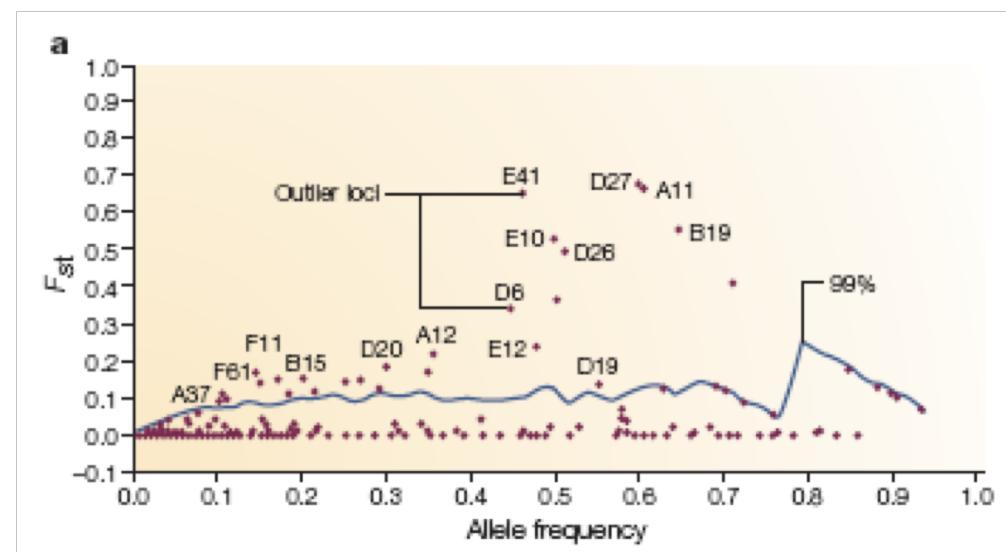


# STEP 3: Conduct statistical tests for outlier loci

- Two general statistical approaches for testing outlier loci:
  1. Theoretical (simulated) null distribution of a summary statistic (e.g.,  $F_{st}$ ):
    - Simulate expected null distribution for neutral loci w/ different population structures ( $Nm$  and  $Ne$ ) and evolutionary histories
  2. Empirical (observed) distribution of summary statistic:
    - Advantage: Controls for demographic history

# ➤ Example of testing for outliers using a theoretical null distribution of $F_{st}$ :

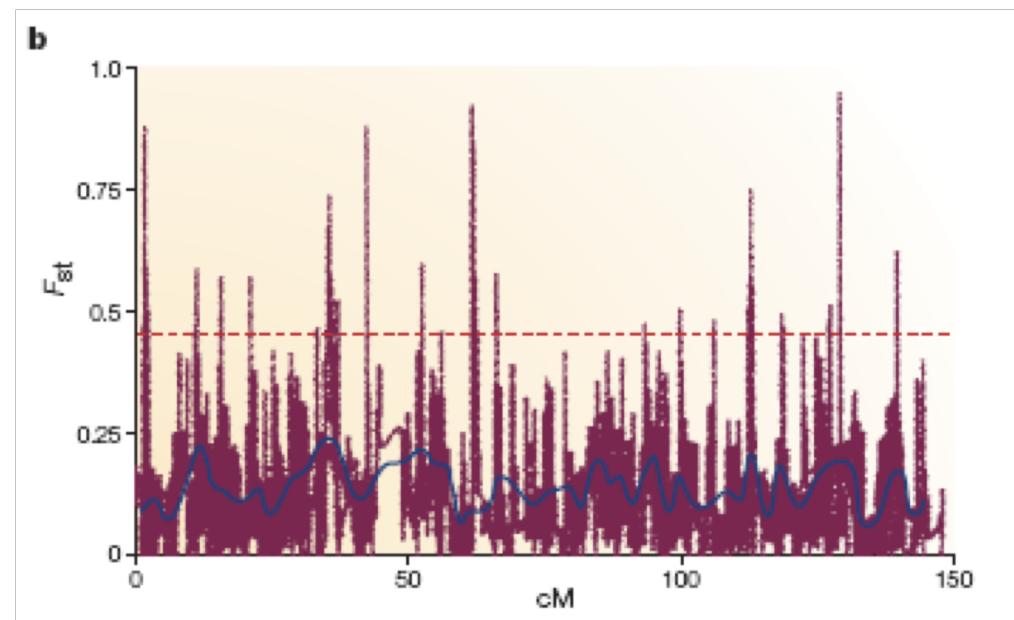
- Wilding et al. (2001) genotyped 306 AFLP loci in the intertidal snail *Littorina saxatilis*
- 15 loci had extremely high  $F_{st}$  values ( $> 0.20-0.30$ ) compared w/ mean of  $<0.04$
- Solid line is the upper 99% CI for simulated neutral loci



➤ Example of testing for outliers using an empirical null distribution of  $F_{st}$ :

- Akey et al. (2002) genotyped many SNP loci on human chromosome 8

- Horizontal dashed line is the threshold for identifying extremely high  $F_{st}$  values



## **STEP 4a: Compute evolutionary or demographic parameters without using outlier loci**

- If interest is in demography/evolutionary history, final step is using validated neutral loci to infer these parameters
  
- This important, because only one or a few outliers can strongly bias estimates

# STEP 4a: Compute evolutionary or demographic parameters without using outlier loci

➤ Example: Effect of outlier loci in *Littorina saxatilis*



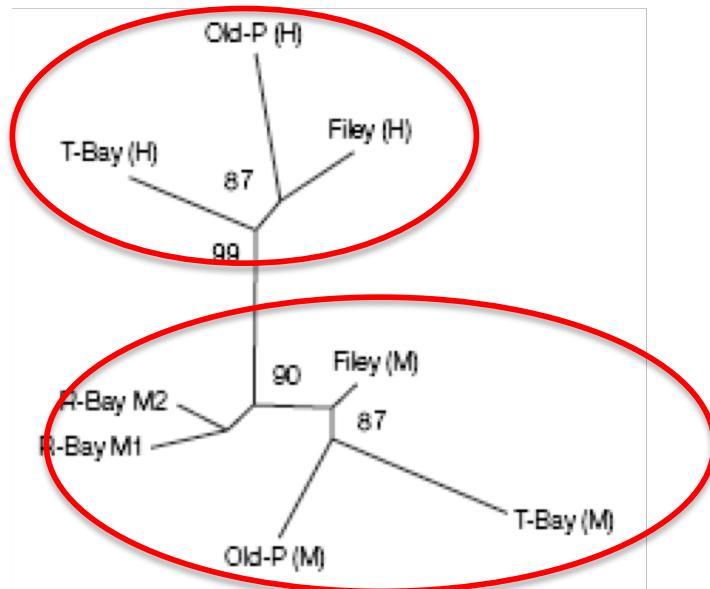
- Two morphotypes:
  1. Thin shell and wide aperture (H) found high in the littoral zone
  2. Thick shell and narrow aperture (M) found low in littoral zone

# STEP 4a: Compute evolutionary or demographic parameters without using outlier loci

➤ Example: Effect of outlier loci in *Littorina saxatilis*

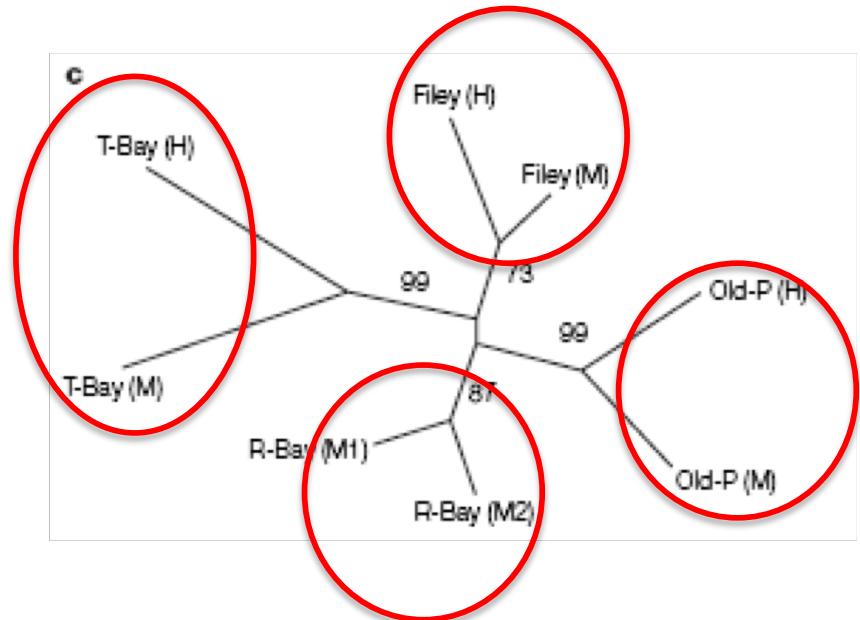
- AFLP phylogeny WITH outlier loci

Groups based on morphotype



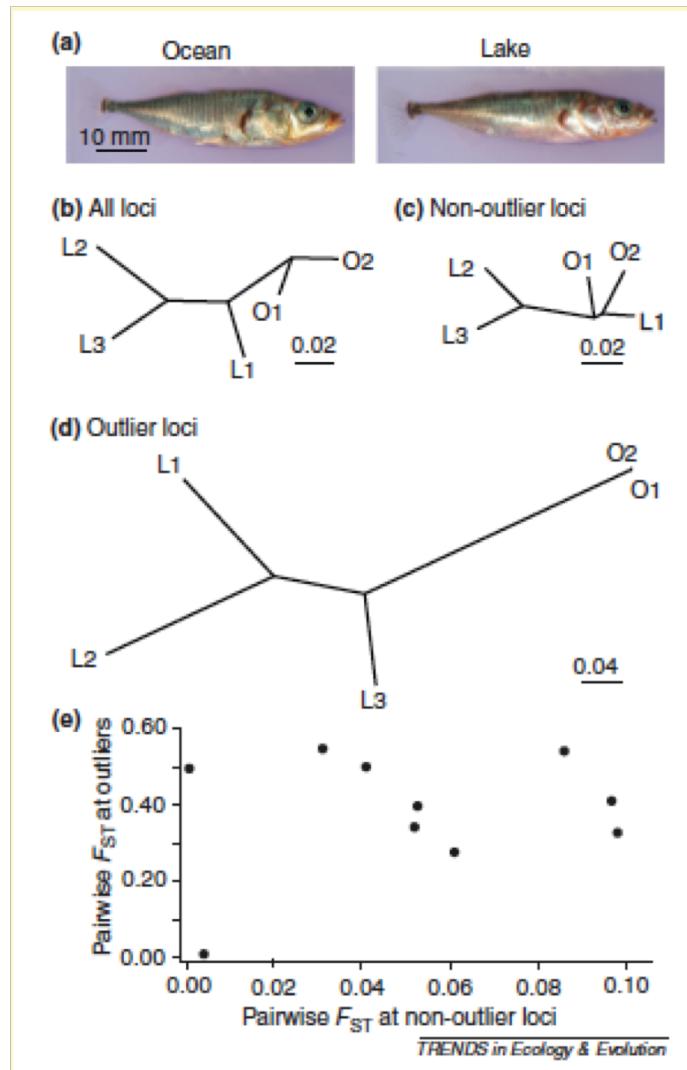
- AFLP phylogeny WITHOUT outlier loci

Groups based on geography



# Harnessing genomics for delineating conservation units

W. Chris Funk<sup>1</sup>, John K. McKay<sup>2</sup>, Paul A. Hohenlohe<sup>3</sup> and Fred W. Allendorf<sup>4</sup>



# STEP 4a: Compute evolutionary or demographic parameters without using outlier loci

Table 2   Some recent studies showing $F_{st}$ -outlier loci and the bias they cause when estimating $N_{em}$										
Species	Number of populations	Number of individuals per population*	Number and type of loci	Number of outlier loci (%)	Mean $F_{st}$ with outlier loci ( $n$ )	Mean $F_{st}$ with non-outlier loci ( $n$ )	$F_{st}$ bias (%)	$N_{em}^{\dagger}$ with all loci	$N_{em}^{\dagger}$ without outlier loci	Refs
<b>Mice</b>										
<i>Peromyscus californicus</i>	13	24	17 alloz	2 (12)	0.367 (17)	0.234 (15)	36	0.43	0.82	26
<i>Peromyscus gossypinus</i>	50	50	37 alloz	2 (5)	0.178 (37)	0.089 (35)	50	1.15	2.56	26
<i>Peromyscus maniculatus</i>	7	60	15 alloz	1 (7)	0.050 (15)	0.019 (14)	62	4.75	12.91	26
<i>Peromyscus leucopus</i>	12	28	33 alloz	3 (9)	0.140 (33)	0.115 (30)	18	1.54	1.92	26
<i>Peromyscus polionotus</i>	28	30	15 alloz	3 (20)	0.382 (15)	0.283 (12)	26	0.40	0.63	26
<b>Sockeye salmon</b>										
<i>Oncorhynchus nerka</i>	4	50	26 alloz, RAPD 1 (4) and msat <sup>§</sup>		0.202 (26)	0.091 (25)	55	0.99	2.50	44
<b>Atlantic cod</b>										
<i>Gadus morhua</i>	6	~100	11 nucl RFLP	1 (9)	0.069 (11)	0.034 (10)	51	3.4	6.6	34,45
<b>Drosophila</b>										
<i>Drosophila melanogaster</i>	15	NA	61 alloz	8 (13)	0.23 (61)	0.17 (53)	26	0.84	1.22	34
<b>Intertidal snails</b>										
<i>Littorina saxatilis</i>	8	50	306 AFLP	15 (5)	0.039**	0.0259**	44**	1.9–3.9††	5.5–308‡‡	6
<b>Humans</b>										
<i>Homo sapiens</i>	3	NA	216 msat	2 (1)	D = 1.34 (216)	D = 0.74 (214)	45	Divergence >70,000 yBP	NA	95
<i>Homo sapiens</i>	3	~40‡‡	8,862 SNP (in genes)	253 (of 8,862) (2.8)%	0.120 = autosomes; 0.195 = X chromosome†††	NA	NA	NA	NA	35

## **STEP 4b: Test for causes of outlier behavior (for example, selection)**

- If interest is in adaptive divergence and selection, final step is to confirm that outlier loci under selection
- Alternative (null) hypothesis is that outlier because of Type I error (erroneous ID)
  - For example, when conducting 100 tests of whether  $F_{ST}$  higher than expected, an average of 5 loci will be erroneously identified as outliers
- Important to have a priori hypotheses to avoid weak inference about cause of selection

## STEP 4b: Test for causes of outlier behavior (for example, selection)

- Ways to confirm outlier behavior caused by selection:
  - Genome position: markers in or near strong candidate genes more likely to be under selection than random markers or markers far from genes
  - Conduct complementary population genomic approaches: for example, Fst outliers within QTLs can show that differentiation b/n ecotypes at adaptive QTL-linked loci maintained by divergent selection
  - Genotype across independent ecological gradients: can provide repeated, independent evidence for outlier behavior

# PROMISE OF POPULATION GENOMICS

- 1. One important approach for identifying loci under selection**
- 2. By IDing loci under selection, can remove these loci and estimate evolutionary and demographic parameters more accurately**
- 3. Outlier loci useful in conservation, for example, for defining “adaptively” distinct populations**

# LIMITATIONS OF POPULATION GENOMICS

1. Significant investment of time and money required for marker development and “genome typing”
2. If just using non-functional markers (e.g., microsatellites, some SNPs), outlier loci do not necessarily = loci underlying local adaptation
  - Therefore, important to use neutral & functional loci

# **POP GEN 101**

- I. What is genetic variation? (terms and basic principles)**
- II. What determines the amount and distribution of genetic variation?**
  - A. Genetic drift**
  - B. Selection**
  - C. Gene flow**
  - D. Mutation**
- III. Effective population size ( $N_e$ )**
- IV. Neutral vs. adaptive genetic variation**
- V. Why genetics is important in conservation: molecular markers as tools & effect of genetics on fitness**
- VI. Population genomics: What is it? Steps involved? Power and limitations.**